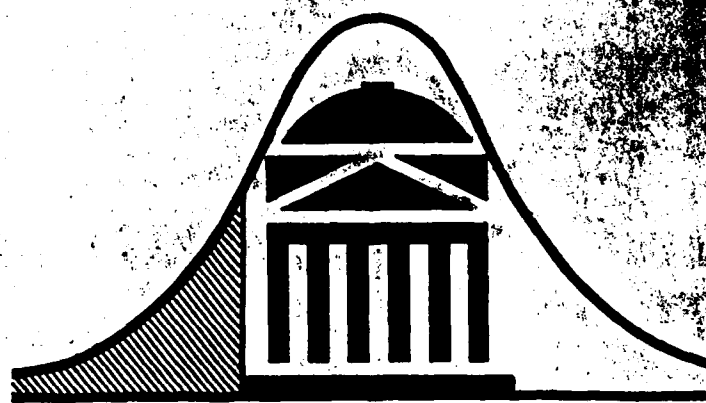


SOUTHERN METHODIST UNIVERSITY

DTIC FILE COPY

AD-A223 280



DTIC
ELECTE
JUN 26 1990
S E D
lo

DEPARTMENT OF STATISTICAL SCIENCE

DALLAS, TEXAS 75275

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

90 06 25 180

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 11 June 1990	3. REPORT TYPE AND DATES COVERED REPRINT		
4. TITLE AND SUBTITLE Saddlepoint Approximations in Conditional Inference		5. FUNDING NUMBERS PE 62714E PR 8A10 TADA WUAH Contract F19628-88-K-0042		
6. AUTHOR(S) Suojin Wang				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Southern Methodist University Dept of Statistical Science Dallas, TX 75275-0332		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Geophysics Laboratory Hanscom AFB, MA 01731-5000		10. SPONSORING/MONITORING AGENCY REPORT NUMBER GL-TR-90-0141		
Contract Manager: James Lewkowicz/LWH				
11. SUPPLEMENTARY NOTES Submitted to the Journal of the Royal Statistical Society, Series B				
12a. DISTRIBUTION/AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED		12b. DISTRIBUTION CODE		
13. ABSTRACT (Maximum 200 words) Saddlepoint approximations are derived for the conditional cumulative distribution function and density of $\bar{X} g(\bar{X}, \bar{Y})$, where (\bar{X}, \bar{Y}) is the sample mean of n iid bivariate random variables and $g(x, y)$ is a nonlinear function of x and y. The relative error of order $O(n^{-1})$ is retained. Such approximations are useful in conditional inference, especially in the case of small or moderate sample sizes. Generalizations to higher dimensional random vectors are also discussed. Two examples are demonstrated and some numerical results are provided to show the extraordinary accuracy of the new approximations.				
14. SUBJECT TERMS Asymptotic expansion Conditional density Conditional distribution		Conditional Inference Nonlinear conditioning Saddlepoint approximation		15. NUMBER OF PAGES 14
17. SECURITY CLASSIFICATION OF REPORT Unclassified		18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified		16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified		18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified		19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified
17. SECURITY CLASSIFICATION OF REPORT Unclassified		18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified		20. LIMITATION OF ABSTRACT SAR

Saddlepoint Approximations in Conditional Inference

Suojin Wang

Department of Statistical Science

Southern Methodist University

Dallas, TX 75275

SUMMARY

Saddlepoint approximations are derived for the conditional cumulative distribution function and density of $\bar{X} \mid g(\bar{X}, \bar{Y})$, where (\bar{X}, \bar{Y}) is the sample mean of n iid bivariate random variables and $g(x, y)$ is a nonlinear function of x and y . The relative error of order $O(n^{-1})$ is retained. Such approximations are useful in conditional inference, especially in the case of small or moderate sample sizes. Generalizations to higher dimensional random vectors are also discussed. Two examples are demonstrated and some numerical results are provided to show the extraordinary accuracy of the new approximations.

Key words: Asymptotic expansion; Conditional density; Conditional distribution; Conditional inference; Nonlinear conditioning; Saddlepoint approximation.

1. INTRODUCTION

Conditional inference plays an important role in statistical inference. The conditionality principle has been used to deal with various problems. Recent developments on this issue include important contributions by Barndorff-Nielsen (1980), Cox and Reid (1987), Fraser and Reid (1988), Diccicco, Field and Fraser (1990) and others. One major step in the procedure of conditional inference is to obtain the conditional distribution functions. As is often the case, the exact conditional distributions are difficult or impossible to obtain, and conventional approximations may often fail to work. For example, generally it is hard to calculate the moments of the conditional distributions which are

✓
necessary quantities for the Edgeworth approximations. Furthermore, these approximations are often unsatisfactory for small or moderate sample sizes.

On the other hand, it is well known that saddlepoint expansions lead to accurate approximations, even for small sample sizes. Among other papers, Barndorff-Nielsen and Cox (1979) and Reid (1988) have discussed the importance and usefulness of the saddlepoint approximations in statistics. Davison and Hinkley (1988) apply the saddlepoint method (see Daniels (1987)) to resampling problems. Using the saddlepoint technique twice, Skovgaard (1987) has derived an accurate approximation to the conditional distribution of a sample mean given a $p-1$ dimensional linear function of a p dimensional variable which has been applied by Davison (1988) and Davison and Hinkley (1988) to generalized linear models and resampling analysis. Wang (1990b) derives saddlepoint formulas for bivariate distribution functions. But it is seen that most applications are limited to the sample mean and other simple linear statistics due to the strong requirements of the saddlepoint technique on the moment generating function of the statistics under consideration.

However, consideration of more general statistics is often desired. In particular, conditional inference often requires a distribution function conditioning on nonlinear functions of a sample mean of multivariate random variables. Recent work on saddlepoint expansions includes expansions for some specific nonlinear statistics by Srivastava and Yau (1989) and Wang (1990c).

In this paper we derive accurate saddlepoint expansions for the case of nonlinear conditioning. The results include Skovgaard's (1987) method as a special case when the distribution is continuous, but have much broader applications. Sections 2 and 3 expand saddlepoint formulas for the conditional density and conditional distribution function, respectively. Two examples are considered in Section 4 to illustrate the use of the new results. Extraordinary accuracy is also shown numerically.

2. SADDLEPOINT EXPANSIONS FOR THE CONDITIONAL DENSITY

We consider first the bivariate case, and generalize results in the $p (> 2)$ dimensional case at the end of Section 3. Assume that $(X_1, Y_1), \dots, (X_n, Y_n)$ are independent identically distributed

bivariate continuous random variables with the cumulant generating function $K(u, v)$ existing in a neighborhood of $(0, 0)$. Let

$$T = c_1 \bar{X} + c_2 \bar{Y}$$

and $Z = g(\bar{X}, \bar{Y})$ be a bivariate function on the sample space, where c_1 and c_2 are given constants, \bar{X} and \bar{Y} be the sample means of X_i 's and Y_i 's, respectively. We wish to expand the conditional distribution $H_n(t_0 | z) = \Pr(T < t_0 | Z = z)$ and the corresponding density. Without loss of generality, let $T = \bar{X}$ and assume that the transformation from (\bar{X}, \bar{Y}) to (T, Z) is one to one. Then the inverse transform can be written as $(\bar{X}, \bar{Y}) = (T, q(T, Z))$ for some function q . When the transform is not one to one, the domain should be partitioned so that in each portion the transform is one to one.

In this section we consider the conditional density. It is well known that the density $p_n(x, y)$ of (\bar{X}, \bar{Y}) has the following saddlepoint approximation (equation (1) of Reid (1988)):

$$p_n(x, y) = p_s(x, y) \{1 + O(n^{-1})\}, \quad (1)$$

where

$$p_s(x, y) = \frac{n}{2\pi \{D(\hat{u}, \hat{v})\}^{1/2}} \exp \{n[K(\hat{u}, \hat{v}) - \hat{u}x - \hat{v}y]\}, \quad (2)$$

$D(u, v) = \{\ddot{K}_{uu} \ddot{K}_{vv} - (\ddot{K}_{uv})^2\} |_{(u,v)} = | \ddot{K}(u, v) |$, the determinant of the 2×2 matrix of the second partial derivatives of K , and (\hat{u}, \hat{v}) is the saddlepoint defined by

$$\begin{cases} \frac{\partial}{\partial u} K(\hat{u}, \hat{v}) = x \\ \frac{\partial}{\partial v} K(\hat{u}, \hat{v}) = y \end{cases} \quad (3)$$

Here we adapt Skovgaard's notations for partial derivatives, e.g., \ddot{K}_{uu} is the second partial derivative of K with respect to u , etc. The relative error in (1) is of order n^{-1} uniformly over compact sets. In this paper we will use " \simeq " to denote an approximation having such an error. Letting $(x, y) = (t, q(t, z))$ and still using (\hat{u}, \hat{v}) to denote the corresponding solution in (3), it is easily seen from (1) that the density $h_n(t, z)$ of (T, Z) can be expressed as

$$h_n(t, z) = p_n(t, q(t, z)) |J| \simeq h_s(t, z), \quad (4)$$

where

$$h_s(t, z) = \frac{n |J|}{2\pi \{D(\hat{u}, \hat{v})\}^{1/2}} \exp \left\{ n[K(\hat{u}, \hat{v}) - \hat{u}t - \hat{u}q(t, z)] \right\},$$

and $J = \dot{q}_z(t, z) = \{\dot{q}_y(x, y)\}^{-1}$ is the Jacobian factor. To approximate the conditional density $h_n(t | z)$ of $T | (Z = z)$, it is sufficient to obtain an approximation to the marginal density $h_n(z)$ for Z . By (4),

$$h_n(z) = \int_{-\infty}^{\infty} h_n(t, z) dt \simeq \int_{-\infty}^{\infty} h_s(t, z) dt. \quad (5)$$

Let

$$L(t) = K(\hat{u}, \hat{v}) - \hat{u}t - \hat{v}q(t, z), \quad (6)$$

where z is fixed so that \hat{u} and \hat{v} depend on t only. Then by (3),

$$L'(t) = -\hat{u} - \hat{v} \dot{q}_t(t, z) \quad (7)$$

and

$$L''(t) = \left\{ -\ddot{K}_{vv}(\hat{u}, \hat{v}) + 2\dot{q}_t(t, z) \ddot{K}_{uv}(\hat{u}, \hat{v}) - [\dot{q}_t(t, z)]^2 \ddot{K}_{uu}(\hat{u}, \hat{v}) \right\} / D(\hat{u}, \hat{v}) - \hat{v} \ddot{q}_{tt}(t, z), \quad (8)$$

since by taking derivative with respect to t of both sides of (3) when $(x, y) = (t, q(t, z))$ we have

$$\begin{cases} \frac{\partial \hat{u}}{\partial t} = \left\{ \ddot{K}_{vv}(\hat{u}, \hat{v}) - \dot{q}_t(t, z) \ddot{K}_{uv}(\hat{u}, \hat{v}) \right\} / D(\hat{u}, \hat{v}) \\ \frac{\partial \hat{v}}{\partial t} = \left\{ -\ddot{K}_{uv}(\hat{u}, \hat{v}) + \dot{q}_t(t, z) \ddot{K}_{uu}(\hat{u}, \hat{v}) \right\} / D(\hat{u}, \hat{v}). \end{cases}$$

We now assume that there exists a unique solution, t_1 , to

$$L'(t) = 0 \quad (9)$$

and $L''(t_1) < 0$. Following the technique by Bleistein (1966), we transform t to w such that

$$(w - w_1)^2 / 2 = L(t_1) - L(t), \quad (10)$$

where $w_1 = \text{sgn}(t_1) \{2[L(t_1) - L(0)]\}^{1/2}$ and $\text{sgn}(w - w_1) = \text{sgn}(t - t_1)$. Note that $w = 0$ and $w = w_1$ correspond to $t = 0$ and $t = t_1$, respectively. From (4), (5) and (10), it is seen that

$$\begin{aligned}
h_n(z) &\simeq \frac{n}{2\pi} \exp\{nL(t_1)\} \int_{-\infty}^{\infty} \exp\left\{-n(w-w_1)^2/2\right\} \frac{|J|}{\{D(\hat{u}, \hat{v})\}^{1/2}} \frac{dt}{dw} dw \\
&\simeq \left\{\frac{n}{2\pi}\right\}^{1/2} \exp\{nL(t_1)\} \left\{ \frac{|J|}{\{D(\hat{u}, \hat{v})\}^{1/2}} \frac{dt}{dw} \right\} \Big|_{w=w_1}.
\end{aligned}$$

This last approximation is obtained by applying Watson's lemma (Jeffreys and Jeffreys (1962)). But by differentiating both sides of (10), we have

$$\left. \frac{dt}{dw} \right|_{w=w_1} = \lim_{t \rightarrow t_1} - \frac{w-w_1}{L'(t)} = \{-L''(t_1)\}^{-1/2}. \quad (11)$$

Thus,

$$h_n(z) \simeq \left\{ -\frac{n}{2\pi L''(t_1) D(\hat{u}_1, \hat{v}_1)} \right\}^{1/2} |J_1| \exp\{nL(t_1)\}, \quad (12)$$

where (\hat{u}_1, \hat{v}_1) is the solution to (3) when $(x, y) = (t_1, q(t_1, z))$, and $J_1 = \dot{q}_z(t_1, z)$ is the Jacobian factor at $t = t_1$. Noticing that the conditional density is given by

$$h_n(t | z) = \frac{h_n(t, z)}{h_n(z)}, \quad (13)$$

from (4), (6) and (12) we have reached the following result.

Theorem 1. Let $T = \bar{X}$, $Z = g(\bar{X}, \bar{Y})$. Assume that the general regularity conditions described at the beginning of this section hold and that the solution t_1 in (9) exists. Denote the inverse transform by $(\bar{X}, \bar{Y}) = (T, q(T, Z))$. Then the conditional density of T given $Z = z$ has the saddlepoint expansion

$$h_n(t | z) \simeq \left\{ \frac{-nL''(t_1) D(\hat{u}_1, \hat{v}_1)}{2\pi D(\hat{u}, \hat{v})} \right\}^{1/2} \left| \frac{J}{J_1} \right| \exp\{n[L(t) - L(t_1)]\}, \quad (14)$$

where (\hat{u}, \hat{v}) and (\hat{u}_1, \hat{v}_1) are the solutions to (3) when $(x, y) = (t, q(t, z))$ and $(t_1, q(t_1, z))$, respectively, and J and J_1 are the Jacobian factor at t and t_1 , respectively.

As is seen in the next section, the technique used to derive (12) can be modified for the purpose of approximating the conditional distribution function. Expansion (12) is useful in the derivations.

3. SADDLEPOINT EXPANSION FOR THE CONDITIONAL DISTRIBUTION

We begin with the following lemma which will be needed in this section.

Lemma 1. Assume that $\alpha(x)$ is an analytic function. Then

$$\left(\frac{n}{2\pi}\right)^{1/2} \int_{-\infty}^y \alpha(x) \exp\{-n x^2/2\} dx \simeq \alpha(0) \Phi(\sqrt{n} y) - \frac{\alpha(y) - \alpha(0)}{\sqrt{n} y} \phi(\sqrt{n} y), \quad (15)$$

where ϕ and Φ are the standard normal density and cumulative distribution function, respectively.

When $y = 0$, the expansion is $\alpha(0)/2 - \alpha'(0)/(2\pi n)^{1/2}$.

This lemma is a special case of Temme (1982); see Temme (1982) for the proof.

By (4) and (6), the conditional distribution function is

$$\begin{aligned} H_n(t_0 | z) &= \frac{1}{h_n(z)} \int_{-\infty}^{t_0} h_n(t, z) dt \\ &\simeq \frac{n}{2\pi h_n(z)} \int_{-\infty}^{t_0} \frac{|J|}{\{D(\hat{u}, \hat{v})\}^{1/2}} \exp\{nL(t)\} dt. \end{aligned} \quad (16)$$

Using the same transformation as in (10) and denoting by w_0 the transformed value of t evaluated at t_0 , we have

$$H_n(t_0 | z) \simeq \frac{n \exp\{nL(t_1)\}}{2\pi h_n(z)} \int_{-\infty}^{w_0} c(w) \exp\{-n(w - w_1)^2/2\} dw,$$

where $c(w) = |J| \frac{dt}{dw} / \{D(\hat{u}, \hat{v})\}^{1/2}$. Therefore by Lemma 1 and formulas (11) and (12), it is easily seen that

$$\begin{aligned}
H_n(t_0 | z) &\simeq \left(\frac{n}{2\pi}\right)^{1/2} \frac{\exp\{nL(t_1)\}}{h_n(z)} \left\{ c(w_1) \Phi(a_0) - \frac{c(w_1) - c(w_0)}{a_0} \phi(a_0) \right\} \\
&\simeq \Phi(a_0) + \left\{ \frac{1}{a_0} - \frac{1}{b_0} \right\} \phi(a_0),
\end{aligned} \tag{17}$$

where

$$\begin{aligned}
a_0 &= \sqrt{n}(w_0 - w_1) = \text{sgn}(t_0 - t_1) \left\{ 2n [L(t_1) - L(t_0)] \right\}^{1/2}, \\
b_0 &= \frac{a_0 c(w_1)}{c(w_0)} = -L'(t_0) \left| \frac{J_1}{J_0} \right| \left\{ -\frac{n D(\hat{u}_0, \hat{v}_0)}{L''(t_1) D(\hat{u}_1, \hat{v}_1)} \right\}^{1/2},
\end{aligned}$$

and J_0 and J_1 are the Jacobian factor at t_0 and t_1 , respectively. We summarize the above results in the following theorem.

Theorem 2. Under the conditions in Theorem 1, the conditional distribution function $H_n(t_0 | z) = \Pr(T < t_0 | Z = z)$ has the saddlepoint expansion (17). When $t_0 = t_1$, the expansion can be obtained by taking the limit of (17) as $t_0 \rightarrow t_1$.

Notice that formula (17) has the same form as Lugannani and Rice's (1980) formula, as well as those of Skovgaard (1987) and Wang (1990c). Note also that because of the special structure of the saddlepoint technique the expansions (4), (12), (14) and (17) with relative error of order n^{-1} are sufficiently accurate for most applications. Higher order expansions could be obtained by referring to Temme (1982), although it is algebraically complicated.

It is straightforward to generalize expansions (14) and (17) to the p (> 2) dimensional case with an analog of Skovgaard (1987). Let (X_i, Y_i) , $i = 1, \dots, n$, be iid continuous p dimensional random vectors. Assume that $(T, Z) = (\bar{X}, g(\bar{X}, \bar{Y}))$ is a one to one transform, where $Z = g(\bar{X}, \bar{Y})$ is a $p-1$ dimensional vector, and that expansions for $H_n(t_0 | z) = \Pr(T < t_0 | Z = z)$ and the density $h_n(t_0 | z)$ are desired. Then the inverse transform $(\bar{X}, \bar{Y}) = (T, q(T, Z))$ exists and the variable v in the cumulant generating function $K(u, v)$ of (\bar{X}, \bar{Y}) is $p-1$ dimensional. Analogously to (6), let

$$L(t) = K(\hat{u}, \hat{v}) - \hat{u}t - \hat{v}^T q(t, z),$$

where \hat{v}^T is the transpose of \hat{v} , and let $J = \frac{\partial(x, y)}{\partial(t, z)}$ be the Jacobian factor and $D(u, v)$ be the determinant of the $p \times p$ matrix of the second partial derivatives of $K(u, v)$. Following the same procedures, it is easily seen that (14) and (17) also hold for the general p dimensional case.

4. EXAMPLES

In this section we consider two applications. The first one is about the conditional density and conditional distribution of $\bar{X} \mid \hat{\sigma}^2$, where $\hat{\sigma}^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$. It is equivalent to consider $\bar{X} \mid \left(\frac{1}{n} \sum X_i^2 - (\bar{X})^2 \right)$. Let $T = \bar{X}$, $\bar{Y} = \frac{1}{n} \sum X_i^2$ and $Z = g(\bar{X}, \bar{Y}) = \bar{Y} - (\bar{X})^2$. Then the problem of approximating $H_n(t_0 \mid z) = \Pr(T < t_0 \mid Z = z)$ and its density $h_n(t \mid z)$ can be handled well by the new results developed in this paper, provided that the general conditions in Section 2 are satisfied. In particular, when X_i 's are normally distributed, say from $N(0, 1)$, it is easily calculated that the saddlepoint expansions (14) and (17) are $h_n(t \mid z) = \sqrt{n} \phi(\sqrt{n} t)$ and $H_n(t_0 \mid z) = \Phi(\sqrt{n} t_0)$, independent of z . That is, the approximations become exact in this case.

We shall discuss the second example in a more detailed manner. In an interesting paper, Hinkley (1977) posed the following problem. Let W_1, \dots, W_n be independently $N(\mu, c^2 \mu^2)$, where c is a known constant and $\mu > 0$ is the parameter of interest. For simplicity, let $c = 1$. We may take as an ancillary statistic

$$Z = \bar{W} / \left\{ \frac{1}{n} \sum W_i^2 \right\}^{1/2}, \quad (18)$$

as pointed out by Hinkley. By the conditionality principle (Cox and Hinkley (1974), Sec. 2.2), inference about μ is based on the conditional distribution of $T = \frac{1}{n} \sum W_i^2$ given the observed value of Z . A major step in the problem is to obtain $H_n(t_0 \mid z)$ and $h_n(t_0 \mid z)$ or sufficiently accurate approximations to them. Expansions (14) and (17) serve this purpose. To illustrate, let the true value $\mu = 1$ and let $X_i = W_i^2$ and $Y_i = W_i$. Thus by (18), $z = g(x, y) = y/\sqrt{x}$ and $y = q(t, z) = \sqrt{t} z$. It is easily obtained that the cumulant generating function of (X_i, Y_i) is

$$K(u, v) = u + v + \frac{(2u + v)^2}{2(1-2u)} - \frac{1}{2} \log(1 - 2u).$$

For $(x, y) = (t, \sqrt{t} z)$, the solution to (3) is

$$(\hat{u}, \hat{v}) = \left(\frac{1}{2} - \frac{1}{2(1-z^2)t}, \frac{\sqrt{t} z}{(1-z^2)t} - 1 \right). \quad (19)$$

From (6), some algebra leads to

$$L(t) = z\sqrt{t} - \frac{t}{2} + \frac{1}{2} \log t + \frac{1}{2} \log(1-z^2).$$

Note that $|z| < 1$. Hence the solution t_1 to $L'(t) = 0$ is given by

$$t_1 = (z^2 + 2 + z\sqrt{z^2+4})/2,$$

and the condition

$$L''(t) = -zt^{-3/2}/4 - t^{-2}/2 < 0$$

holds for all t in its domain $(0, \infty)$. Moreover, one can obtain that

$$D(\hat{u}, \hat{v}) = 2\{(1-z^2)t\}^3 \text{ and } |J| = \sqrt{t}.$$

Therefore, expansion (14) reduces to

$$h_n(t | z) \simeq c_n t^{n/2-1} \exp \{-n(\sqrt{t}-z)^2/2\}, \quad (20)$$

where

$$c_n = \frac{1}{2} \left\{ \frac{n}{\pi} \left(\frac{z\sqrt{t_1}}{2} + 1 \right) \right\}^{1/2} t_1^{-n/2} \exp \{n(\sqrt{t_1}-z)^2/2\}.$$

Since the exact density $h_n(t | z)$ is proportional to $t^{n/2-1} \exp\{-n(\sqrt{t}-z)^2/2\}$ (see Hinkley (1977)),

(20) is viewed as "exact" if the renormalization is allowed. Using the above calculations, expansion

(17) for $H_n(t_0 | z)$ is easily computed with

$$a_0 = \text{sgn}(t_0 - t_1) \left\{ n \left[(\sqrt{t_0} - \sqrt{t_1})(\sqrt{t_0} + \sqrt{t_1} - 2z) + \log(t_1/t_0) \right] \right\}^{1/2},$$

$$b_0 = (t_0 - z\sqrt{t_0} - 1) \left\{ n/(z\sqrt{t_1} + 2) \right\}^{1/2}.$$

Table 1 compares the saddlepoint approximation (17) with the true conditional distribution function for $z = 0.5$ and $n = 5, 10$. The extraordinary accuracy of the approximation is evident uniformly over the domain of the statistic. Note that calculations show that the relative error is not affected by different values of z , i.e., the numerical accuracy of the expansion (17) is stable for all $|z| < 1$.

5. CONCLUDING REMARKS

In this paper we have derived genuine saddlepoint expansions for conditional densities and distributions with nonlinear conditioning. The resulting formulas are highly accurate even for small sample sizes and are easy to use. Formula (17) reduces to Skovgaard's (1987) formula when the conditioning variable $Z = g(\bar{X}, \bar{Y})$ is a linear function of \bar{X} and \bar{Y} .

Resampling methods such as bootstrap are very useful in conditional inference (see, for example, Hinkley (1988)). However, computational difficulty is generally a major obstacle in such approaches, evidently more so than in classical parametric problems. Besides their direct use in parametric applications, the new approximations provide a possible remedy to overcome the difficulties in many such cases. Further investigations are needed to draw a solid conclusion on this issue. We have assumed continuity of the statistics in our derivations while those in the resampling schemes are discrete, although they become smooth as the sample size increases. There is now an interesting question as to whether the formulas derived here or their modified versions are valid for such purposes. A detailed examination of this question is beyond the scope of this paper. Recent work by Wang (1990a) on the validity of saddlepoint expansions in the problem of bootstrapping a sample mean is relevant.

ACKNOWLEDGEMENT

The author wishes to thank Professor D.V. Hinkley for initially raising the problem and Professor H.L. Gray for his encouragement on this research. It was partially supported by DARPA/AFGL Contract No. F19628-88-K-0042.

REFERENCES

- Barndorff-Nielsen, O. (1980). Conditional resolutions. *Biometrika*, 67, 293-310.
- Barndorff-Nielsen, O. and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. R. Statist. Soc. B*, 41, 279-312.
- Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B*, 49, 1-39.
- Daniels, H.E. (1987). Tail probability approximations. *Int. Statist. Rev.* 55, 37-48.
- Davison, A.C. (1988). Approximate conditional inference in generalized linear models. *J. R. Statist. Soc. B*, 50, 445-461.
- Davison, A.C. and Hinkley, D.V. (1988). Saddlepoint approximations in resampling methods. *Biometrika*, 75, 417-431.
- Diciccio, T.J., Field, C.A. and Fraser, D.A.S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika*, 77, 77-95.
- Fraser, D.A.S. and Reid, N. (1988). On conditional inference for a real parameter: A differential approach on the sample space. *Biometrika*, 75, 251-264.
- Hinkley, D.V. (1977). Conditional inference about a normal mean with known coefficient of variation. *Biometrika*, 64, 105-108.
- Hinkley, D.V. (1988). Bootstrap methods (with discussion). *J.R. Statist. Soc. B*, 50, 321-337.
- Jeffreys, H. and Jeffreys, B.S. (1962). *Methods of Mathematical Physics*, 3rd ed. London and New York: Cambridge University Press.

- Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.*, 12, 475-490.
- Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.*, 3, 213-238.
- Skovgaard, I.M. (1987). Saddlepoint expansions for conditional distributions. *J. Appl. Prob.*, 24, 875-887.
- Srivastava, M.S. and Yau, W.K. (1989). Tail probability approximations of a general statistic. Unpublished manuscript.
- Temme, N.M. (1982). The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions. *SIAM J. Math. Anal.*, 13, 239-251.
- Wang, S. (1990a). Saddlepoint approximations in resampling analysis. *Ann. Inst. Statist. Math.*, 42, to appear.
- Wang, S. (1990b). Saddlepoint approximations for bivariate distributions. *J. Appl. Prob.*, 27, to appear.
- Wang, S. (1990c). Saddlepoint approximations for nonlinear statistics. *Canadian J. Statist.*, 18, to appear.

Table 1. Saddlepoint approximation (17) to the conditional distribution $H_n(t_0 | z)$
in the second example in Section 4; $z = 0.5$.

n = 5			n = 10		
t_0	saddlepoint	true	t_0	saddlepoint	true
0.03	.00004	.00004	0.15	.00002	.00002
0.05	.00016	.00016	0.20	.00011	.00011
0.10	.00103	.00102	0.40	.00351	.00351
0.30	.01803	.01803	0.60	.02220	.02223
0.50	.06295	.06306	0.80	.07111	.07119
0.70	.13418	.13446	1.00	.15644	.15660
1.00	.27298	.27354	1.20	.27179	.27202
1.40	.47269	.47348	1.50	.46870	.46900
1.80	.64674	.64754	1.80	.65067	.65097
2.20	.77748	.77815	2.10	.79010	.79034
2.80	.89815	.89857	2.40	.88318	.88335
3.30	.95012	.95036	2.80	.95160	.95169
4.00	.98293	.98303	3.30	.98588	.98591
5.50	.99860	.99861	4.20	.99884	.99884
6.50	.99976	.99977	4.70	.99974	.99974
7.50	.99996	.99996	5.40	.99997	.99997